PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2003-271616**
(43)Date of publication of application : **26.09.2003**

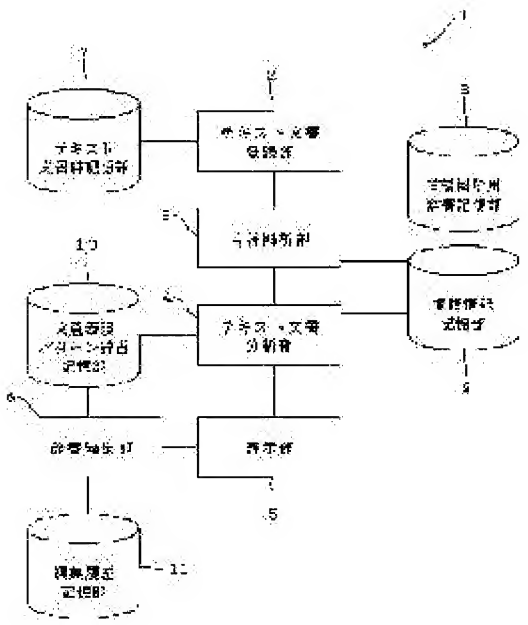
(51)Int.Cl. G06F 17/30

(21)Application number : **2002-068858**
(22)Date of filing : **13.03.2002**

(71)Applicant : **RICOH CO LTD**
(72)Inventor : **SATO NAOKO**

(54) DOCUMENT CLASSIFICATION DEVICE, DOCUMENT CLASSIFICATION METHOD AND RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a document classification device, document classification method and recoding medium for classifying a text document on the basis of a desired meaning.
SOLUTION: In this document classification device 1, a language analysis part 3 retrieves a text document including an inputted retrieval request from a plurality of text documents collected and accumulated in a text document group storage part 2, and classifies the plurality of text documents by categories on the basis of the text document retrieval result. A text document analysis part 4 classifies, on the basis of the category classification result, the text documents by meanings in reference to the meaning expression pattern dictionary of a meaning expression pattern dictionary storage part 10 in which characteristic meaning expression patterns for expressing meanings are registered, and a dictionary editing part 6 optionally edits the meaning expression pattern dictionary according to a user's operation. Accordingly, the classification reference can be optionally changed to perform the document classification according to the meaning desired by the user, and the convenience can be improved.



LEGAL STATUS

[Date of request for examination] 01.03.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2003－271616
（P2003－271616A）
(43)公開日　平成15年９月26日(2003.9.26)

(51)Int.Cl.[7]	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ０６Ｆ　17/30	２１０	Ｃ０６Ｆ　17/30	２１０Ｄ　５Ｂ０７５
	１７０		１７０Ａ
	２２０		２２０Ｚ

審査請求　未請求　請求項の数７　ＯＬ　（全 11 頁）

(21)出願番号　特願2002－68858(P2002－68858)

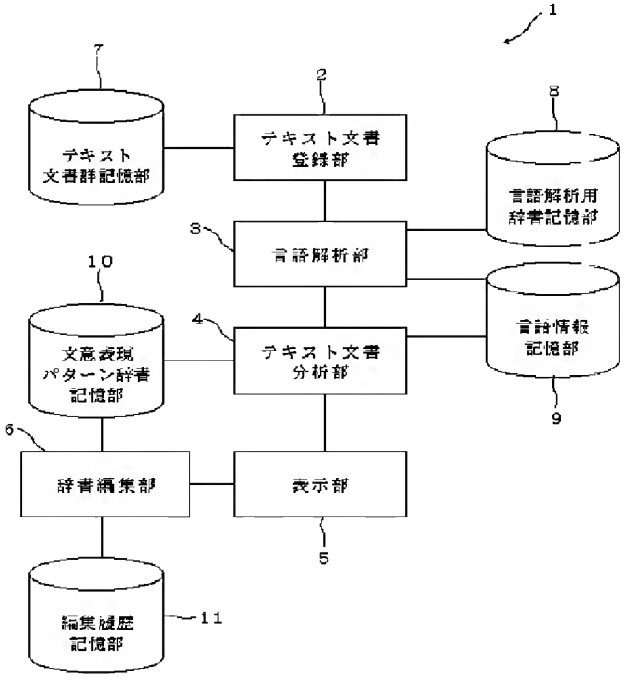
(22)出願日　平成14年３月13日(2002.3.13)

(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(72)発明者　佐藤　奈穂子
東京都大田区中馬込１丁目３番６号　株式会社リコー内

Ｆターム(参考)　5B075 ND03 ND34 NK06 NK32 NK39
NR02 NR12 NS10 UU06

(54)【発明の名称】　文書分類装置、文書分類方法及び記録媒体

(57)【要約】
【課題】本発明はテキスト文書を所望の文意による文書分類する文書分類装置、文書分類方法及び記録媒体を提供する。
【解決手段】文書分類装置１は、言語解析部３で、入力された検索要求を含むテキスト文書を、テキスト文書群記憶部２に収集・蓄積された複数のテキスト文書から検索し、言語解析部３で、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、テキスト文書分析部４が、当該カテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書記憶部１０の文意表現パターン辞書を参照してテキスト文書を文意別に分類し、文意表現パターン辞書を、ユーザの操作に応じて、辞書編集部６で、任意に編集する。したがって、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、利用性を向上させることができる。



【特許請求の範囲】
【請求項１】入力された検索要求を含むテキスト文書を文書蓄積手段に収集・蓄積された複数のテキスト文書から検索するテキスト文書検索手段と、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成手段と、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を記憶する文意表現パターン辞書記憶手段と、前記カテゴリ生成手段のカテゴリ分け結果に基づいて前記文意表現パターン辞書を参照して前記テキスト文書を文意別に分析・分類する文意分類手段と、を備えた文書分類装置であって、前記文意表現パターン辞書記憶手段の文意表現パターン辞書を任意に編集する文意表現パターン辞書編集手段を備えていることを特徴とする文書分類装置。
【請求項２】前記文意表現パターン辞書は、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集手段は、前記文意表現パターン辞書に、前記意図タグと前記文意表現パターンを新たに追加登録することを特徴とする請求項１記載の文書分類装置。
【請求項３】前記文意表現パターン辞書は、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集手段は、前記文意表現パターン辞書に登録されている前記意図タグと前記文意表現パターンを個々に無効化することを特徴とする請求項１または請求項２記載の文書分類装置。
【請求項４】文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法において、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行うことを特徴とする文書分類方法。
【請求項５】前記文書分類方法は、前記文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集処理ステップが、前記文意表現パターン辞書に、前記意図タグと前記文意表現パターンを新たに追加登録することを特徴とする請求項４記載の文書分類方法。
【請求項６】前記文書分類方法は、前記文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集処理ステップが、前記文意表現パターン辞書に登録されている前記意図タグと前記文意表現パターンを個々に無効化することを特徴とする請求項１または請求項２記載の文書分類方法。
【請求項７】文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法のプログラム及びデータを記憶する記録媒体であって、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行う文書分類方法のプログラム及びデータを記録することを特徴とする記録媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、文書分類装置、文書分類方法及び記録媒体に関し、詳細には、テキスト文書をカテゴリ分けし、文意表現パターン辞書を参照して文意別に分析・分類するとともに、文意表現パターン辞書を任意に編集し、所望の文意による文書分類を行う文書分類装置、文書分類方法及び当該文書分類方法のプログラムとデータを記録した記録媒体に関する。
【０００２】
【従来の技術】近時、情報の電子化が進み、従来紙文書で保管されていた文書も電子化されるようになってきている。このような文書の電子化に伴って、大量の電子化文書が流通し、収集・蓄積された電子化文書をいかに管理して簡便に再利用するかが重量な問題となってきている。
【０００３】そして、従来、このような電子文書の簡便な利用や管理のためにさまざまな文書処理技術が提案されており、この文書処理技術の一例として、ある目的で収集された文書群の自動分類が挙げられる。この文書群の自動分類は、大量の電子化文書群から類似した文書群を自動分類する技術であり、一般的には、各文書に含まれている重要語句の類似性、出現頻度、出現場所等の共通点に基づいて、関連性の高い文書をグルーピングする仕組みになっている。この文書群の自動分類で利用されている重要語句としては、従来、文書におけるキーワードが用いられており、主に文書中に頻出する名詞、動詞等の品詞を限定して抽出した単語である。

【０００４】また、従来、処理対象となる文書を入力する入力部と、入力した文書中の文章に対して形態素解析を行う形態素解析部と、前記形態素解析部から出力された形態素列の部分列を、重み付きで特定表現候補として取得する特定表現候補取得部と、予めいくつかの特定表現を格納した特定表現辞書と、与えられた形態素列の前記特定表現辞書中の表現に対するマッチ度を表す実数を、当該形態素列の前記特定表現辞書に対する検索結果として取得する特定表現辞書検索部と、前記特定表現候補に対して、前記候補に付与された重みと、前記特定表現検索部による前記候補の前記特定表現辞書に対する検索結果とを変数として判別スコアを計算し、前記判別スコアが予め設定した一定の値を下回る候補を除外する判別分析実行部と、前記特定表現候補のうち、前記判別分析実行部によって除外されなかった形態素の文字列を特定表現として出力する出力部とを備えた文書処理装置が提案されている（特開２００１－７５９５９号公報参照）。

【０００５】すなわち、この従来技術は、人名や企業名等に特有の特定表現を、辞書を用いて抽出し、それを重要語句として、単語単体よりも、複数の単語の並びや単語の出現パターンなど含有情報が多く、より限定がかかった表現単位での類似性判断を行なうことで、分類の精度の向上を図っている。

【０００６】

【発明が解決しようとする課題】しかしながら、上記公報記載の従来技術は、人名や企業名等の特定表現の情報抽出を目的としたものであり、特定表現を利用した分類処理については、言及されておらず、また、特定表現パターンの内容として、書き手の文意を表わす表現パターンについては、上記公報の実施例に記載されていない。

【０００７】一方、アンケートの自由記述部分などを分類、分析する場合、設問に関する話題は予め分かっており、書き手の意図にこそ分析のポイントがあると考えられ、書き手の文意を表わす表現パターンが要求されることは必至である。また、その際に、意図表現パターンをユーザが目的に応じて解釈し、分類（分析）基準を任意に変更できるようなしくみが要望されている。

【０００８】そこで、請求項１記載の発明は、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、入力された検索要求を含むテキスト文書を、テキスト文書検索手段で、文書蓄積手段に収集・蓄積された複数のテキスト文書から検索し、カテゴリ生成手段で、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、カテゴリ生成手段のカテゴリ分け結果に基づいて、文意分類手段が、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照してテキスト文書を文意別に分析・分類し、当該文意表現パターン辞書を、文意表現パターン辞書編集手段で、任意に編集することにより、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行い、利用性の良好な文書分類装置を提供することを目的としている。

【０００９】請求項２記載の発明は、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録することにより、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行い、利用性の良好な文書分類装置を提供することを目的としている。

【００１０】請求項３記載の発明は、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化することにより、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにし、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性の良好な文書分類装置を提供することを目的としている。

【００１１】請求項４記載の発明は、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、テキスト文書検索処理ステップで、入力された検索要求を含むテキスト文書を文書蓄積手段のテキスト文書から検索し、カテゴリ生成処理ステップで、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、文意分類処理ステップで、カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書を任意に編集することにより、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行い、利用性の良好な文書分類方法を提供することを目的としている。

【００１２】請求項５記載の発明は、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録することにより、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行い、利用性の良好な文書分類方法を提供することを目的としている。

【００１３】請求項６記載の発明は、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文

書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化することにより、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにし、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性の良好な文書分類方法を提供することを目的としている。

【００１４】請求項７記載の発明は、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法のプログラム及びデータを記憶する記録媒体に、当該プログラム及びデータとして、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行う文書分類方法のプログラム及びデータを記録することにより、当該プログラム及びデータをコンピュータ等の情報処理装置に導入することで、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行い、利用性を向上させることのできる文書分類装置及び文書分類方法を実現する記録媒体を提供することを目的としている。

【００１５】

【課題を解決するための手段】請求項１記載の発明の文書分類装置は、入力された検索要求を含むテキスト文書を文書蓄積手段に収集・蓄積された複数のテキスト文書から検索するテキスト文書検索手段と、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成手段と、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を記憶する文意表現パターン辞書記憶手段と、前記カテゴリ生成手段のカテゴリ分け結果に基づいて前記文意表現パターン辞書を参照して前記テキスト文書を文意別に分析・分類する文意分類手段と、を備えた文書分類装置であって、前記文意表現パターン辞書記憶手段の文意表現パターン辞書を任意に編集する文意表現パターン辞書編集手段を備えていることにより、上記目的を達成している。

【００１６】上記構成によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、入力された検索要求を含むテキスト文書を、テキスト文書検索手段で、文書蓄積手段に収集・蓄積された複数のテキスト文書から検索し、カテゴリ生成手段で、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、カテゴリ生成手段のカテゴリ分け結果に基づいて、文意分類手段が、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照してテキスト文書を文意別に分析・分類し、当該文意表現パターン辞書を、文意表現パターン辞書編集手段で、任意に編集するので、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、利用性を向上させることができる。

【００１７】この場合、例えば、請求項２に記載するように、前記文意表現パターン辞書は、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集手段は、前記文意表現パターン辞書に、前記意図タグと前記文意表現パターンを新たに追加登録するものであってもよい。

【００１８】上記構成によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録するので、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行うことができ、利用性を向上させることができる。

【００１９】また、例えば、請求項３に記載するように、前記文意表現パターン辞書は、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集手段は、前記文意表現パターン辞書に登録されている前記意図タグと前記文意表現パターンを個々に無効化するものであってもよい。

【００２０】上記構成によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化するので、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにすることができ、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性を向上させることができる。

【００２１】請求項４記載の発明の文書分類方法は、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法において、入力された検索要

求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行うことにより、上記目的を達成している。

【００２２】上記構成によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、テキスト文書検索処理ステップで、入力された検索要求を含むテキスト文書を文書蓄積手段のテキスト文書から検索し、カテゴリ生成処理ステップで、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、文意分類処理ステップで、カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書を任意に編集するので、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、利用性を向上させることができる。

【００２３】この場合、例えば、請求項５に記載するように、前記文書分類方法は、前記文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集処理ステップが、前記文意表現パターン辞書に、前記意図タグと前記文意表現パターンを新たに追加登録してもよい。

【００２４】上記構成によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録するので、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行うことができ、利用性を向上させることができる。

【００２５】また、例えば、請求項６に記載するように、前記文書分類方法は、前記文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとが対として登録されており、前記文意表現パターン辞書編集処理ステップが、前記文意表現パターン辞書に登録されている前記意図タグと前記文意表現パターンを個々に無効化してもよい。

【００２６】上記構成によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化するので、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにすることができ、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性を向上させることができる。

【００２７】請求項７記載の発明の記録媒体は、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法のプログラム及びデータを記憶する記録媒体であって、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行う文書分類方法のプログラム及びデータを記録することにより、上記目的を達成している。

【００２８】上記構成によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類する文書分類方法のプログラム及びデータを記憶する記録媒体に、当該プログラム及びデータとして、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行う文書分類方法のプログラム及びデータを記録しているので、当該プログラム及びデータをコンピュータ等の情報処理装置に導入することで、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行い、利用性を向上させることのできる文書分類装置及び文書分類方法を実現することができる。

【００２９】

【発明の実施の形態】以下、本発明の好適な実施の形態を添付図面に基づいて詳細に説明する。なお、以下に述べる実施の形態は、本発明の好適な実施の形態であるから、技術的に好ましい種々の限定が付されているが、本発明の範囲は、以下の説明において特に本発明を限定する旨の記載がない限り、これらの態様に限られるものではない。

【００３０】図１～図４は、本発明の文書分類装置、文書分類方法及び記録媒体の一実施の形態を示す図であり、図１は、本発明の文書分類装置、文書分類方法及び記録媒体の一実施の形態を適用した文書分類装置１のブロック構成図である。

【００３１】図１において、文書分類装置１は、テキスト文書登録部２、言語解析部３、テキスト文書分析部４、表示部５、辞書編集部６、テキスト文書群記憶部７、言語解析用辞書記憶部８、言語情報記憶部９、文意表現パターン辞書記憶部１０及び編集履歴記憶部１１等を備えている。文書分類装置１は、文書分類処理プログラム及び必要なデータを記録するＣＤ－ＲＯＭ（Compact Disc Read Only Memory）等の記録媒体を、例えば、コンピュータ等に読み取らせて導入することで、構築される。

【００３２】テキスト文書群記憶部（文書蓄積手段）７は、収集されたテキスト文書のテキスト文書群が登録され、登録されたテキスト文書が文書分類・分析の対象となる。

【００３３】テキスト文書登録部２は、収集されたテキスト文書をテキスト文書群記憶部７に登録して蓄積させ、登録したテキスト文書群の管理を行う。

【００３４】言語解析用辞書記憶部８は、言語解析部３による言語解析に必要な各種言語解析情報を記憶する。

【００３５】言語情報記憶部９は、言語解析部３によるテキスト文書の解析処理によって得られる言語的属性を解析単位毎に記憶する。

【００３６】言語解析部（テキスト文書検索手段、カテゴリ生成手段）３は、入力された検索要求を含むテキスト文書をテキスト文書群記憶部７に収集・蓄積された複数のテキスト文書から検索するテキスト文書検索処理と、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理を行う。具体的には、言語解析部３は、言語解析用辞書記憶部８の記憶する言語解析用辞書に基づいて、テキスト文書登録部２によりテキスト文書群記憶部７に登録された各テキスト文書の形態素を解析する形態素解析処理、形態素解析処理の解析結果と文節生成規則に基づいて文節を生成する文節生成処理、文節生成処理で生成した文節が解析対象のテキスト文書のどの文節に係っているかを予め設定されている係り受け解析規則に基づいて特定する係り受け解析処理等の各ステップ処理を実行し、これらの形態素解析処理、文節生成処理、係り受け解析処理によって得られる言語的属性を解析単位に言語情報記憶部９に保持させる言語情報登録処理を行う。

【００３７】文意表現パターン辞書記憶部（文意表現パターン辞書記憶手段）１０は、文意を表現する特徴的な文意表現パターン辞書が登録され、蓄積する。この文意表現パターン辞書記憶部１０に登録される文意表現パターン辞書としては、例えば、図２に示すようなものであり、文書の書き手の意図を示す文意タグの種類（例えば、依頼、要望、否定、疑問、推量等）毎に、文意表現パターンがそれぞれ対として複数登録されている。例えば、依頼の文意タグに対しては、「用言＋て／欲しい」、「用言＋て／ください」、「サ変名詞＋ください」等の文書表現パターン、要望の文意タグに対しては、「用言＋助動詞たい」、「用言＋て／いただき・助動詞たい」、「サ変名詞＋いただき・助動詞たい」等の文書表現パターン、否定の文意タグに対しては、「用言＋助動詞ない」、「サ変名詞＋助動詞＋助動詞ない」等の文書表現パターン、疑問の文意タグに対しては、「用言＋終助詞か」、「文末＝記号？」等の文書表現パターン、推量の文意タグに対しては、「用言＋助動詞だろう」、「用言＋副助詞か＋副助詞も＋しれない」、「用言＋助動詞そう」、「助言＋助動詞できる＋助動詞そう」等の文書表現パターンが対として登録されている。

【００３８】辞書編集部（文意表現パターン辞書編集手段）６は、文意表現パターン辞書記憶部１０に登録されている文意表現パターン辞書を、ユーザの操作に応じて、任意に編集し、例えば、文意タグと文意表現パターンを文意表現パターン辞書に新たに追加登録し、また、文意表現パターン辞書に記載されている文書の書き手の意図を示す文意タグと、それを特定するための特徴的な言語表現を個々に無効化する。

【００３９】編集履歴記憶部１１は、辞書編集部６による文意表現パターン辞書記憶部１０の文意表現パターン辞書の編集履歴を記憶する。

【００４０】テキスト文書分析部（文書分類手段）４は、言語解析部３による解析結果の各解析単位または各単位における言語的属性に基づいて、テキスト文書から文意表現パターン辞書記憶部１０の文意表現パターン辞書を検索可能な単位を生成し、この文意表現パターン辞書を検索可能な単位として、例えば、各テキスト文書に対して、文節を生成する文意別文節処理を行う。また、テキスト文書分析部４は、日本語では、文末に意図表現が表出するという特性を利用するために、文末文節を抽出する。

【００４１】すなわち、テキスト文書分析部４は、言語解析部３の解析結果である各解析単位または各単位における言語属性を、文書表現パターン辞書記憶部８の文書表現パターンと文意タグを検索可能な形式に変換する文意表現変換処理、当該文意表現変換処理した言語属性に基づいて文書表現パターン辞書記憶部８の辞書引きを行

う辞書引き処理、当該辞書引き処理で辞書引きした文書表現パターンに合致した文書表現パターンを文意タグに変換する文意タグ変換処理及び文意タグ変換処理で変換した文意タグを用いてテキスト文書の文意別カテゴリの分類を行う文意別カテゴリ分類処理を行う。

【００４２】表示部５は、液晶ディスプレイやＣＲＴ（陰極線管：Cathode Ray Tube）等が用いられ、テキスト文書分析部４の分析したり、分類した結果を表示し、また、辞書編集部６の編集結果等を表示する。

【００４３】次に、本実施の形態の作用を説明する。文書分類装置１は、文書分類処理プログラム及び必要なデータを記録するＣＤ－ＲＯＭ等の記録媒体を、例えば、コンピュータ等の情報処理装置に読み取らせて導入することで、構築され、電子化されたテキスト文書群を言語解析して、書き手の意図を表現する表現パターンをテキスト文書中から重要語句として抽出し、文書を分類するところにその特徴がある。

【００４４】すなわち、文書分類装置１は、分析対象のテキスト文書群が入力されると、当該テキスト文書群をテキスト文書登録部２が当該テキスト文書群をテキスト文書群記憶部５に登録する。

【００４５】いま、例えば、あるマリンスポーツについて意見を収集・蓄積した、下記のようなテキストデータがあり、集めた意見を文意別に分類・分析するものとして、以下説明する。

～ 収集テキスト ～

１．お金がかかりそう。
２．サーフィンをやれる環境をもっと良くしてほしい！
３．もっと盛んになってほしい。
４．全てのコトが忘れられて、すごく楽しそう。
５．安く手軽にできるならやってみたい。
６．リフレッシュできそう。
７．もっと海岸でのマナーを大切に指導して欲しい。
８．もっと安くして欲しい。
９．とても楽しそうだが面倒くさそう。
１０．お金がいっぱいかかるけど楽しそう。
１１．一回やったらハマリそう。

【００４６】まず、最初に、これらのテキスト文書に対して、言語解析部３で言語解析を行い、テキストの構成単語の品詞等の属性情報を取得する。この言語解析は、既存のさまざまな手法で実現することができる。

【００４７】次に、テキスト文書分析部４が、言語解析部３の解析結果に基づいて、文意表現パターン辞書記憶部１０の文意表現パターン辞書を検索可能な単位を生成、例えば、各テキストに対して、文節を単位として、生成する。この文節生成技術は、既存の言語処理技術で実現することができる。

【００４８】次に、テキスト文書分析部４は、日本語では文末に意図表現が表出するという特性を利用して、各テキスト文書について、以下のように、文末文節を抽出する。

１．かかり・そう・。
２．良く・し・て・ほしい・！
３．なっ・て・ほしい・。
４．楽し・そう・。
５．やっ・て・み・たい・。
６．リフレッシュ・でき・そう・。
７．指導・し・て・欲しい・。
８．し・て・欲しい・。
９．面倒くさ・そう・。
１０．楽し・そう・。
１１．ハマリ・そう・。

【００４９】テキスト文書分析部４は、この単位で、品詞等の属性を用いて、さらに正規表現に変換して記憶し、これらに対して、図２に示した文意表現パターン辞書を辞書引きする。この辞書引きにより、上記例では、以下のような辞書引き結果を得ることができる。

１．用言＋助動詞そう	文意（推量）
２．用言＋て／ほしい	文意（依頼）
３．用言＋て／ほしい	文意（依頼）
４．用言＋助動詞そう	文意（推量）
５．用言＋助動詞たい	文意（要望）
６．用言＋助動詞できる＋助動詞そう	文意（推量）
７．用言＋て／欲しい	文意（依頼）
８．用言＋て／欲しい	文意（依頼）
９．用言＋助動詞そう	文意（推量）
１０．用言＋助動詞そう	文意（推量）
１１．用言＋助動詞そう	文意（推量）

この辞書引きの結果、文意（推量）は、｛１，４，６，９，１０，１１｝、文意（依頼）は、｛２，３，７，８｝、文意（要望）は、｛５｝という文書の文意別カテゴリ分類を行うことができる。

【００５０】そして、文書分類装置１は、上記文書分類結果を得たユーザが、上記テキスト文書の例の場合に、文意（依頼）は文意（要望）に併せたほうが適切であると考える場合、辞書編集機能を利用して、文意（依頼）の文意表現パターンの無効化や文意（要望）の文意表現パターンの新規追加を行なうことができる。

【００５１】すなわち、ユーザが文書分類装置１の図示しない操作部で、ボタンやコマンド入力等を行って、辞書編集の指示を出すと、文書分類装置１は、表示部５の表示画面に、例えば、図３に示すような辞書編集ウインドウを表示させる。文書分類装置１は、この辞書編集ウインドウとして、図３に示すように、文意表現パターン辞書記憶部１０の文意表現パターン辞書に登録されている文意と文意表現パターン対を表示し、また、各文意表現パターン対に、無効化欄が設けられている。

【００５２】ユーザは、この文意表現パターン対の無効化欄に、図３に示すように、チェックを入れることで、当該文意の対として登録されている文意表現パターンを個別に無効化することができ、文書分類装置１は、無効化欄にチェックの入れられた文意の対として登録されている文意表現パターンを個別に無効化する。

【００５３】さらに、文書分類装置１は、辞書編集ウインドウに設けられている新規追加ボタン（図３の右下に示されている新規追加のボタン）が押されると、追加ウインドウを開き、既存の文意の呼び出しや新規の文意の登録の操作を可能とする。この追加ウインドウで、例えば、新規ボタンが押されると、その文意タグと対になる文意表現パターンの新規登録を行う。

【００５４】例えば、図３では、上記テキスト文書の例において、文意（依頼）に対で登録されていた３パターンを無効化し、追加ウインドウで既存文意（要望）を呼び出し、追加ウインドウで呼び出した既存文意（要望）に、先に無効化した文意（依頼）に登録されていた３パターンを登録する処理が示されている。

【００５５】文書分類装置１は、ユーザの所望する文意表現パターン辞書の辞書編集が終わると、辞書編集部６が、当該編集結果を文意表現パターン辞書記憶部１０の文意表現パターン辞書に保存し、これまでの辞書内容と編集後の辞書内容との辞書の差異を編集履歴として編集履歴記憶部１１に保存する。

【００５６】辞書編集部６は、ユーザの操作部からの編集履歴参照操作に応じて、編集履歴記憶部１１に記憶されている編集履歴を、表示部５に表示し、ユーザが閲覧できるようにする。

【００５７】ユーザは、編集履歴を参照して、選択した履歴を、文意表現パターン辞書に反映させて利用することもできる。

【００５８】このように編集した辞書を用いて、上記テキスト文書の例において、再度分類を行なうと、文意（依頼）だったデータが、文意（要望）と認識され、分類結果は、文意（推量）が｛１，４，６，９，１０，１１｝、文意（要望）が｛２，３，５，７，８｝という文書の文意別カテゴリ分類が実現される。

【００５９】すなわち、図４に示すように、辞書編集部６は、辞書編集指示があると（ステップＳ１０１）、辞書編集ウインドウを起動して、表示部５に表示し（ステップＳ１０２）、無効化欄にチェックがあるか判別する（ステップＳ１０３）。

【００６０】ステップＳ１０３で、無効化欄にチェックがないときには、新規追加指示があるか判別し（ステップＳ１０４）、新規追加指示がないときには、ステップＳ１０１に戻って上記同様に処理する。

【００６１】ステップＳ１０３で、無効化欄にチェックがあると、辞書編集部６は、無効化欄にチェックの入っている文意の対として登録されている文意表現パターンを個別に無効化する無効化処理を行い（ステップＳ１０５）、新規追加指示があるかチェックする（ステップＳ１０４）。

【００６２】ステップＳ１０４で、新規追加指示があると、辞書編集部６は、追加ウインドウを起動して表示部５に表示させ、既存文意が選択されたかチェックする（ステップＳ１０７）。

【００６３】ステップＳ１０７で、既存文意が選択されると、辞書編集部６は、当該選択された文意を文意表現パターン辞書記憶部１０の文意表現パターン辞書から呼び出し（ステップＳ１０８）、文意表現パターン辞書に登録する（ステップＳ１０９）。

【００６４】ステップＳ１０７で、既存文意が選択されないときには、辞書編集部６は、新規文意の登録であると判断して、新規文意を文意表現パターン辞書に登録する（ステップＳ１０９）。

【００６５】文意表現パターン辞書への登録を行うと、辞書編集部６は、保存指示があるかチェックし（ステップＳ１１１）、保存指示があると、当該文意表現パターン辞書の編集履歴を編集履歴記憶部１１に保存して、処理を終了する（ステップＳ１１２）。

【００６６】このように、本実施の形態の文書分類装置１は、言語解析部３で、入力された検索要求を含むテキスト文書を、テキスト文書群記憶部２に収集・蓄積された複数のテキスト文書から検索し、言語解析部３で、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、テキスト文書分析部４が、当該カテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書記憶部１０の文意表現パターン辞書を参照してテキスト文書を文意別に分析・分類し、当該文意表現パターン辞書を、辞書編集部６で、任意に編集している。

【００６７】したがって、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、文書分類装置１の利用性を向上させることができる。

【００６８】また、本実施の形態の文書分類装置１は、文意表現パターン辞書記憶部１０の文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、辞書編集部６が、ユーザの操作に応

じて、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録している。

【００６９】したがって、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行うことができ、利用性を向上させることができる。

【００７０】さらに、本実施の形態の文書分類装置１は、文意表現パターン辞書記憶部１０の文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、辞書編集部６が、ユーザの操作に応じて、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化している。

【００７１】したがって、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにすることができ、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性を向上させることができる。

【００７２】以上、本発明者によってなされた発明を好適な実施の形態に基づき具体的に説明したが、本発明は上記のものに限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【００７３】

【発明の効果】請求項１記載の発明の文書分類装置によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、入力された検索要求を含むテキスト文書を、テキスト文書検索手段で、文書蓄積手段に収集・蓄積された複数のテキスト文書から検索し、カテゴリ生成手段で、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、カテゴリ生成手段のカテゴリ分け結果に基づいて、文意分類手段が、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照してテキスト文書を文意別に分析・分類し、当該文意表現パターン辞書を、文意表現パターン辞書編集手段で、任意に編集するので、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、利用性を向上させることができる。

【００７４】請求項２記載の発明の文書分類装置によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録するので、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行うことができ、利用性を向上させることができる。

【００７５】請求項３記載の発明の文書分類装置によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集手段が、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化するので、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにすることができ、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性を向上させることができる。

【００７６】請求項４記載の発明の文書分類方法によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を分析・分類するに際して、テキスト文書検索処理ステップで、入力された検索要求を含むテキスト文書を文書蓄積手段のテキスト文書から検索し、カテゴリ生成処理ステップで、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けして、文意分類処理ステップで、カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書を任意に編集するので、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行うことができ、利用性を向上させることができる。

【００７７】請求項５記載の発明の文書分類方法によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に、意図タグと文意表現パターンを新たに追加登録するので、文意表現パターン辞書に未登録の文意や対応する文意表現パターンを適宜登録して、幅広い文意分類を行うことができ、利用性を向上させることができる。

【００７８】請求項６記載の発明の文書分類方法によれば、文意表現パターン辞書に、文書の書き手の意図を示す意図タグと、当該文書の書き手の意図を特定する特徴的な文意表現パターンとを対として登録し、文意表現パターン辞書編集処理ステップで、文意表現パターン辞書に登録されている意図タグと文意表現パターンを個々に無効化するので、文意表現パターンを別文意へ再編成したり、文意表現パターンを削除することなく、無効化とすることで、ユーザが無効化を解除したり、過去履歴を利用する際に再利用できるようにすることができ、ユーザの所望する文意により、より一層適切に文書分類を行って、より一層利用性を向上させることができる。

【００７９】請求項７記載の発明の記録媒体によれば、文書蓄積手段に収集・蓄積された複数のテキスト文書を

分析・分類する文書分類方法のプログラム及びデータを記憶する記録媒体に、当該プログラム及びデータとして、入力された検索要求を含むテキスト文書を前記文書蓄積手段のテキスト文書から検索するテキスト文書検索処理ステップと、当該テキスト文書検索結果に基づいて複数のテキスト文書をカテゴリ分けするカテゴリ生成処理ステップと、前記カテゴリ生成処理ステップでのカテゴリ分け結果に基づいて、文意を表現する特徴的な文意表現パターンの登録されている文意表現パターン辞書を参照して前記テキスト文書を文意別に分類する文意分類処理ステップと、前記文意表現パターン辞書を任意に編集する文意表現パターン辞書編集処理ステップと、の各ステップ処理を行う文書分類方法のプログラム及びデータを記録しているので、当該プログラム及びデータをコンピュータ等の情報処理装置に導入することで、分類基準を任意に変更して、ユーザの所望する文意により文書分類を行い、利用性を向上させることのできる文書分類装置及び文書分類方法を実現することができる。

【図面の簡単な説明】

【図１】本発明の文書分類装置、文書分類方法及び記録媒体の一実施の形態を適用した文書分類装置の要部ブロック構成図。

【図２】図１の文書表現パターン辞書記憶部に格納されている文意表現パターン辞書の一例を示す図。

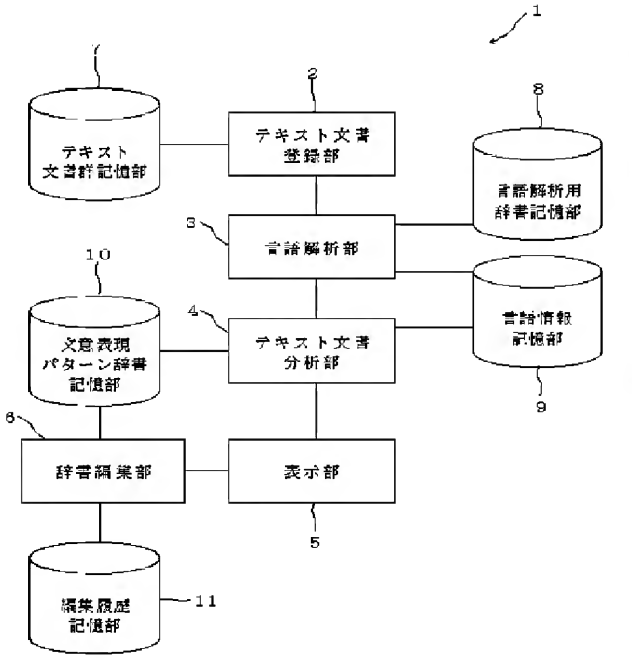
【図３】図１の表示部に表示される辞書編集ウインドウの一例を示す図。

【図４】図１の文書分類装置による文意表現パターン辞書編集処理を示すフローチャート。

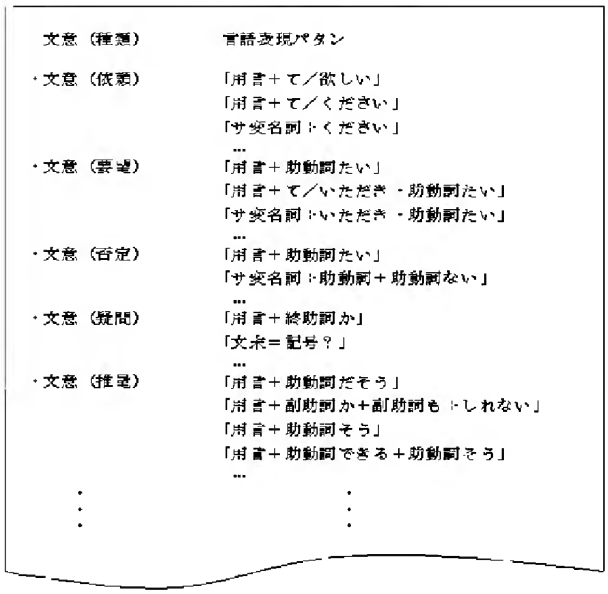
【符号の説明】

１　文書分類装置
２　テキスト文書登録部
３　言語解析部
４　テキスト文書分析部
５　表示部
６　辞書編集部
７　テキスト文書群記憶部
８　言語解析用辞書記憶部
９　言語情報記憶部
１０　文意表現パターン辞書記憶部
１１　編集履歴記憶部

【図１】

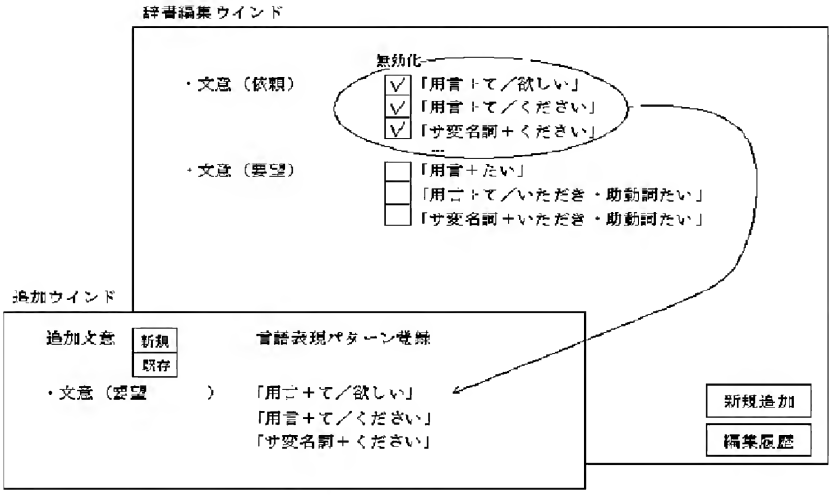


【図２】

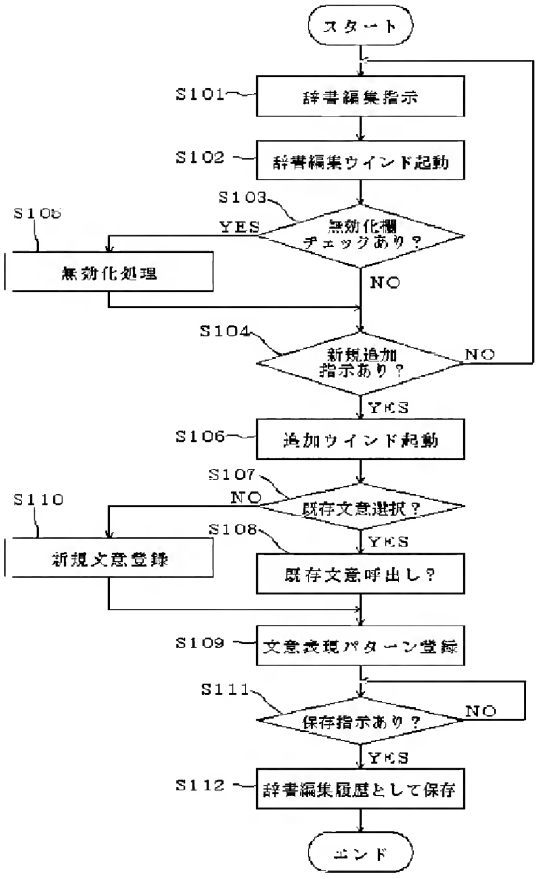


文意（種類）	言語表現パタン
・文意（依頼）	「用言＋て／欲しい」 「用言＋て／ください」 「サ変名詞＋ください」 …
・文意（要望）	「用言＋助動詞たい」 「用言＋て／いただき・助動詞たい」 「サ変名詞＋いただき・助動詞たい」 …
・文意（否定）	「用言＋助動詞たい」 「サ変名詞＋助動詞＋助動詞ない」 …
・文意（疑問）	「用言＋終助詞か」 「文末＝記号？」 …
・文意（推量）	「用言＋助動詞だそう」 「用言＋副助詞か＋副助詞も＋しれない」 「用言＋助動詞そう」 「用言＋助動詞できる＋助動詞そう」 …
・ ・ ・	・ ・ ・

【図3】


辞書編集ウインド
無効化
・文意（依頼）
「用言＋て／欲しい」
「用言＋て／ください」
「サ変名詞＋ください」
・文意（要望）
「用言＋たい」
「用言＋て／いただき・助動詞たい」
「サ変名詞＋いただき・助動詞たい」
追加ウインド
追加文意
新規
既存
言語表現パターン登録
・文意（要望　　）
「用言＋て／欲しい」
「用言＋て／ください」
「サ変名詞＋ください」
新規追加
編集履歴

【図4】


スタート
S101
辞書編集指示
S102
辞書編集ウインド起動
S103
無効化欄チェックあり？
YES
S105
無効化処理
NO
S104
新規追加指示あり？
NO
YES
S106
追加ウインド起動
S107
既存文意選択？
NO
S110
新規文意登録
YES
S108
既存文意呼出し？
S109
文意表現パターン登録
S111
保存指示あり？
NO
YES
S112
辞書編集履歴として保存
エンド